Supplemental Information

Supplemental Figure Legends

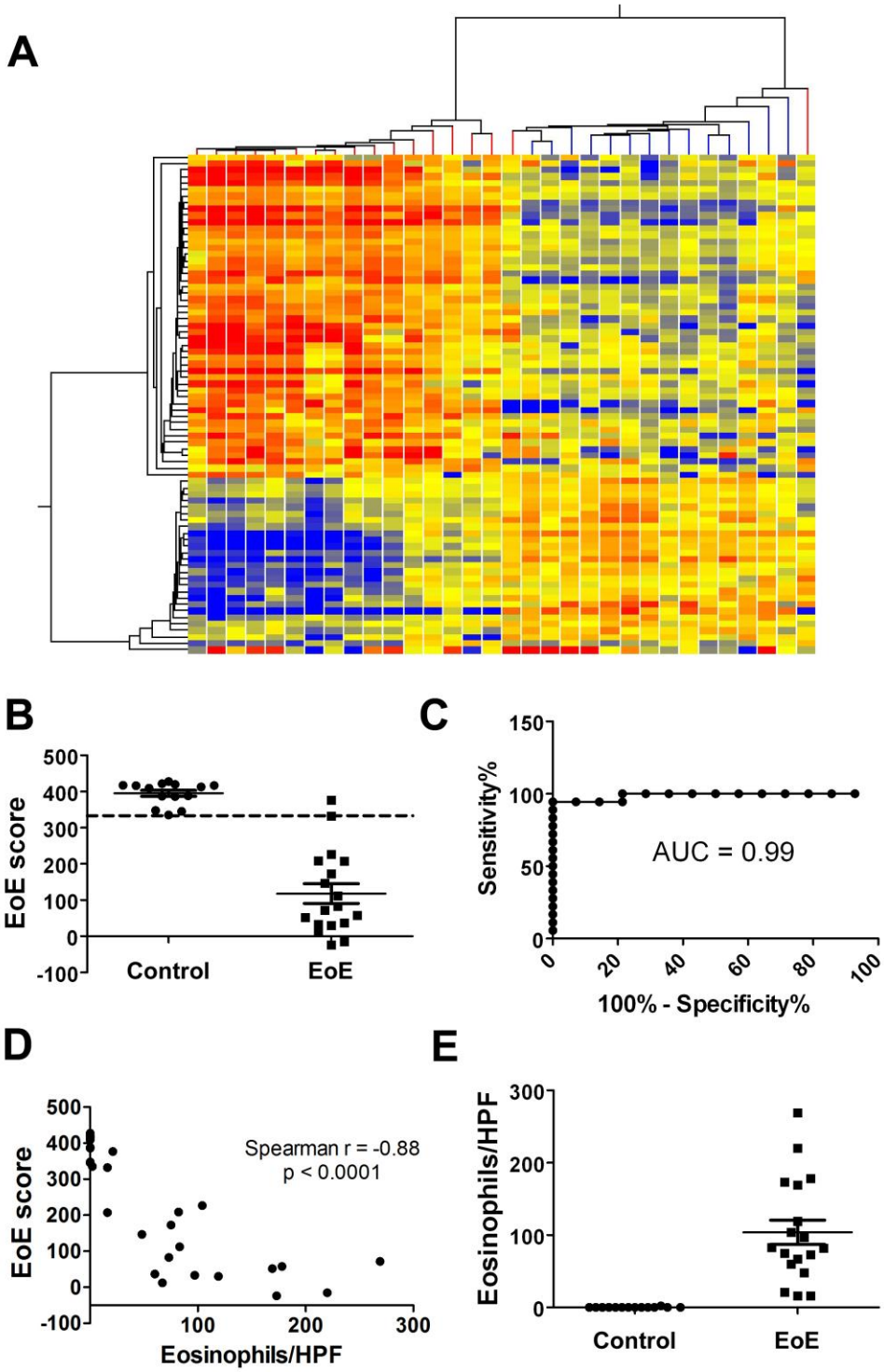

Figure S1. Replication study with independent cohorts of control and EoE. A replication study was performed on 14 control patients and 18 patients with eosinophilic esophagitis (EoE). Histological counts and the 15 eosinophils / HPF cut-off were used as the standard for EoE herein. A, The double-clustered heat map from 14 control (blue branch) patients and 18 patients with EoE (red branch) is shown with the first branch of the x-axis tree indicating EoE diagnosis. B, Using the EoE score cut-off of 333 (dashed line), control patients and patients with EoE are well distinguished by the EoE score algorithm. C, The diagnostic merit for these 32 patients was assessed by ROC analysis with an area under the curve (AUC) of 0.99. D, The negative correlation between EoE score and tissue eosinophilia was assayed by regression analysis with Spearman r and p values shown. E, Eosinphils/HPF distribution among the 32 patients studied herein, control vs. EoE. All scatter plots were graphed as mean ± SEM.

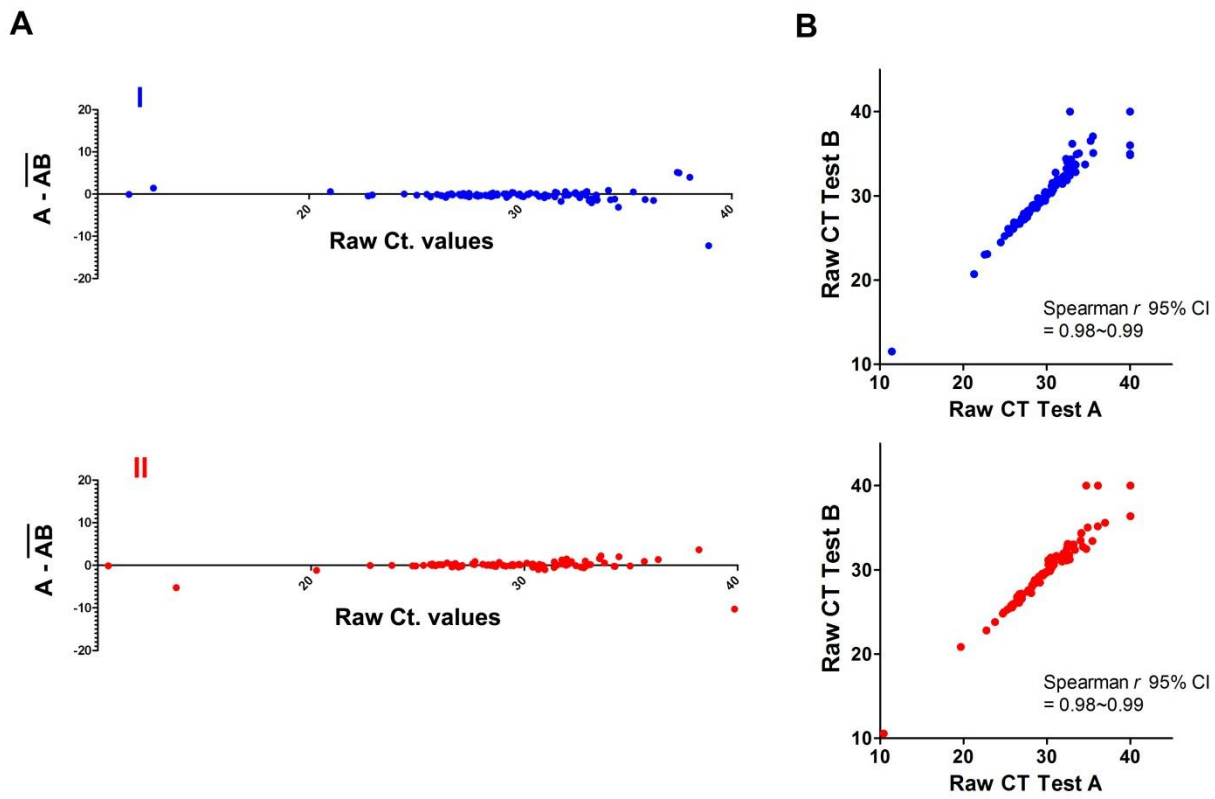

Figure S2. Reproducibility of EDP analysis. A, We assessed the EDP reproducibility by reverse-transcribing and PCR amplifying two RNA samples (I and II) with a different fluidic card half a year apart (data A and data B). By Bland-Altman analysis, the repetitive sets of expression data from fresh sample I (blue) and II (red) were plotted on raw CT value of 96 genes (x-axis) vs. the difference between raw CT value A and the average of CT value A and B (AB bar), reflecting the variance between the two data sets. B, Raw CT values on each of the 96 genes from sample I (blue) and II (red) were double plotted between the two repetitive tests (Test A vs. Test B) and subjected to correlation analysis. Spearman r with 95% CI is indicated on the chart.

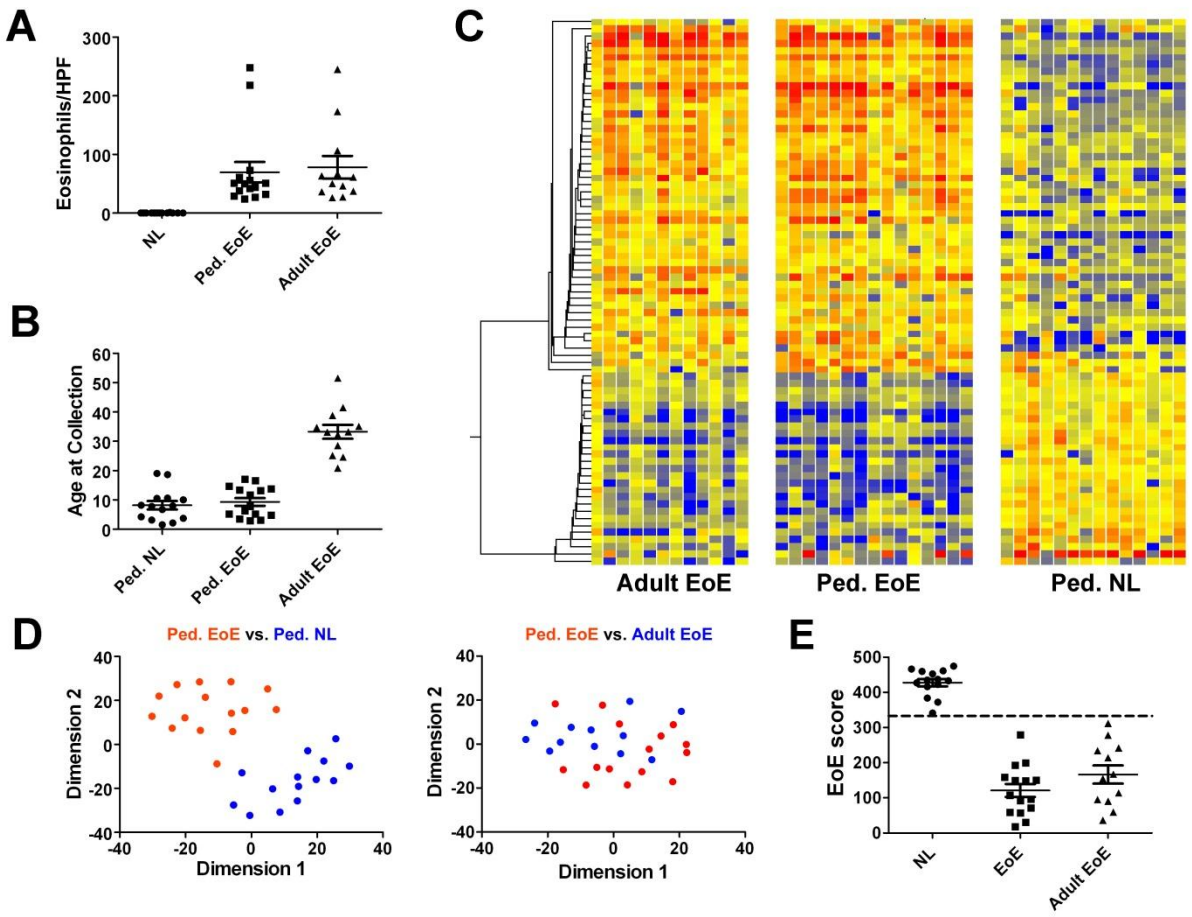

Figure S3. Adult and pediatric EoE share comparable transcriptomes. A, Eosinophils/HPF were graphed for all 3 groups to show the eosinophilia match in the two active EoE cohorts. B, The age distribution chart for all 3 groups analyzed herein. C, Molecular expression heat maps from eosinophilia-matched adult patients with active eosinophilic esophagitis (EoE) were acquired by EDP and juxtaposed to the heat maps of pediatric (Ped.) EoE and pediatric normal (NL) cohorts for expression signature comparison (n = 12). D, Left panel: expression profile of 77 core EoE genes from pediatric EoE (red) and NL (blue) cohorts were reduced to 2-D visualization by MDS analysis, with distance between two given points reflecting expression difference in 77-D space (Euclidean metric). Right panel: MDS analysis was performed between pediatric EoE (red) and adult EoE (blue) cohorts with 2-D data plotted on the same scale. E, EoE scores were calculated for adult EoE, pediatric EoE, and pediatric NL cohorts to compare the EoE signature in 1-D quantification. All scatter plots were graphed as mean ± SEM.

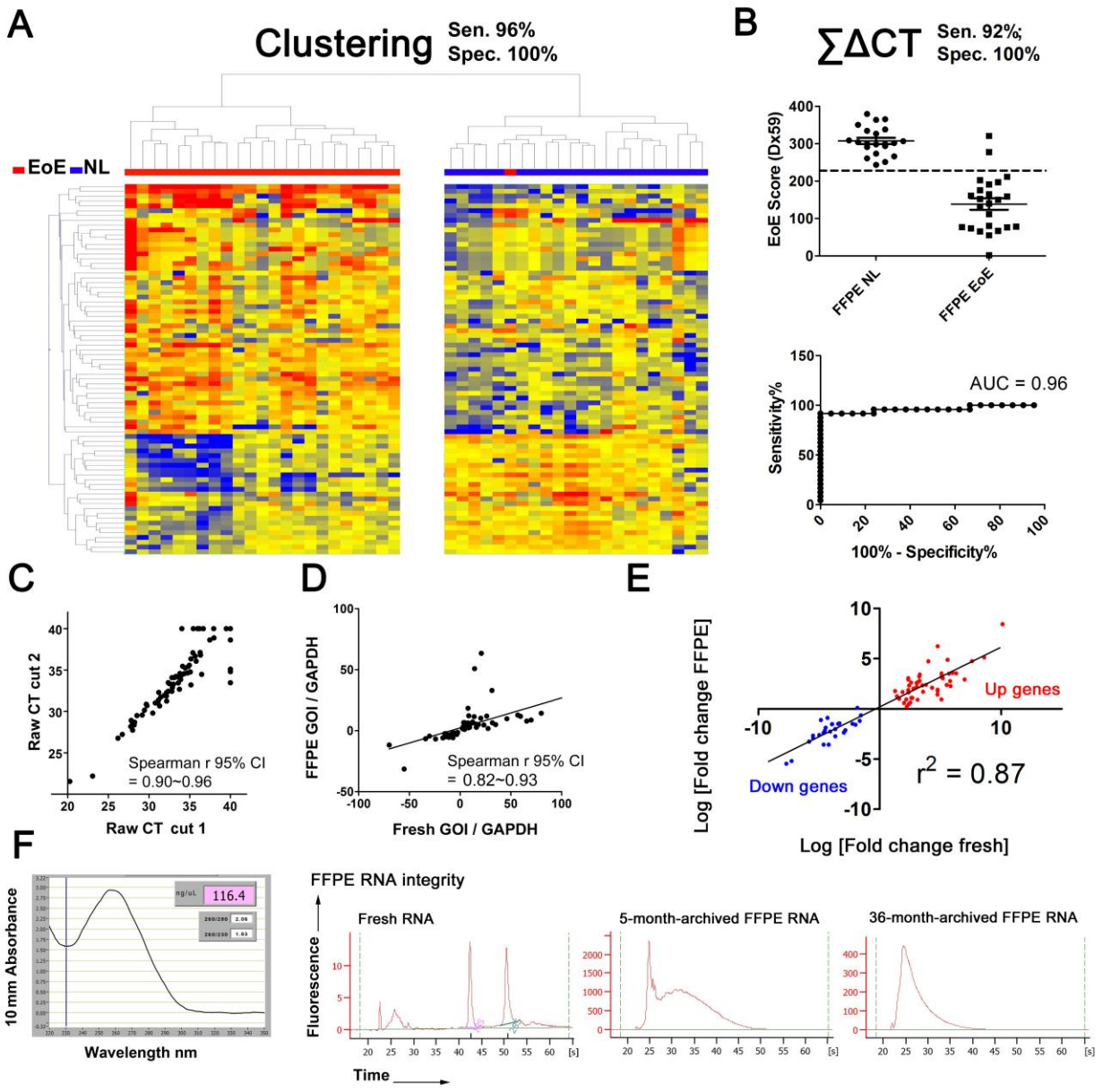

Figure S4. The FFPE sample diagnosing capacity of EDP with both algorithms. A, Eosinophilic esophagitis (EoE, n = 24) and normal (NL, n = 21) formalin-embedded, paraffin-fixed **(**FFPE) samples were subjected to RNA extraction from 60 μm of paraffin sections from a single biopsy. The heat diagram of all 45 FFPE samples based on the previously mentioned 77 core EoE genes was generated with 2-D clustering. The cluster algorithm discrimination is indicated on the top bar of the heat map (EoE, red; NL, blue). B, As to FFPE diagnosis with EoE score algorithm (ΣΔCT), a EoE score of 59 significant genes (FFPE EoE vs. NL [total n=45], corrected $p < 0.05$, fold change > 2.0) was calculated and plotted, with the corresponding ROC curve exhibited in the lower panel. A cut-off of 355 was derived from ROC analysis optimizing both sensitivity (Sen.) and specificity (Spec.). The grouped scattered plots were graphed as mean ± SEM. C, To evaluate the reproducibility for sequential FFPE sample sections of different but

adjacent tissue input, EDP amplification was independently performed from different sections on the same tissue block. The regression analysis between raw CT values of cut 1 and cut 2 was graphed with Spearman r 95% CI displayed. D, A EDP gene expression correlation study between a pair of concurrent FFPE and fresh biopsies from the same patient was shown (same visit, but different biopsies from disparate distal esophageal loci). The regression graph between fresh and FFPE individual gene of interest (GOI) expressions (normalized to GAPDH) was shown with Spearman r 95% CI displayed. E, With the scope of 77 core diagnostic genes, collective dysregulation vector (bi-directional expression fold change, EoE vs. NL) correlation between algorithm developing fresh cohorts (n = 29) and the above FFPE cohorts (n = 45) was plotted with $r^2 = 0.87$, $p < 0.0001$. F, Representative RNA quality from 80 μm FFPE tissue was shown with high purity and abundance enough for qPCR array (left panel). A series of representative FFPE RNA integrity assessments by Agilent Bioanalyser analysis is shown (right panel). The archiving period is noted, and fresh RNA control is shown.

Supplemental Materials and Methods

Bioinformatics selection of representative EoE genes

The selection was performed based on a set of mRNA microarray data expanded from the 2006 Blanchard et al. publication [1] and the 2010 Caldwell et al. publication [2]. Notably, a number of these EoE genes has been validated by biochemical means [1, 3-7]. After a comprehensive expression profile analysis of ~1000 EoE genes by Genespring GX 11.5 software (Agilent Technologies), the selection of the 94 EoE-representative genes was performed based on the following considerations: dysregulation (p-value and fold-change), bi-directional dysregulation coverage, capacity to predict steroid exposure (glucocorticoid-responding genes), capacity to differentiate EoE remission from NL and fibrosis indicators, Th2 cytokine expression profile, biological significance, functional clustering (supplemental Table S1 end column), FFPE compatibility and RNA degradation direction, splice variants and fluidic card design format In addition, since some of the cytokine-producing cells are too scarce to be detected by the microarray despite playing a vital role in EoE pathogenesis, we also included biomarkers for several immunocytes that are known to affect allergic responses (see Table 1 for a representative illustration of EoE gene categories and Table S1 for a categorized list of all 96 genes in the initial design; see Table S2 for glucocorticoid-remission-related genes).

Additional sample inclusion and IRB regulation

All of the samples analyzed herein in this study do not overlap with the initial microarray study by Blanchard et al.[1]. In addition, no sample is included in more than one sub-study performed herein, except that the overall study of 166 samples was a collection of all fresh RNA samples involved, 26 impedance pH study FFPE samples were used to collectively evaluate the FFPE

diagnostic merit, and the discovery NL and EoE cohorts were used as reference cohorts in some sub-studies as mentioned therein.

The distal esophageal biopsy was used throughout the study, as this is typically obtained during endoscopy, represents the conventional location of biopsies, and readily allows comparison to the control samples, including GERD, which is generally limited to this region.

This study was approved by the Institutional Review Board (IRB) of the Cincinnati Children's Hospital Medical Center (2008-0090). Written informed consent was received from participants prior to inclusion in the study. The entire study has been conducted according to Declaration of Helsinki principles.

Diagnosis algorithm development

The raw amplification data were acquired/analyzed by RQ manager 2.0 (Applied Biosystems) and then exported into Genespring GX 11.5 for further relative expression analysis and algorithm development. Non-amplified reaction was assigned a CT value of 40. A false discovery rate (FDR)-corrected , two-tailed student T-test was performed between the 14 NL and 15 EoE algorithm-developing patients simultaneously with a 2-fold change filter, which resulted in 77 significant genes (FDR-corrected $p < 0.05$) dysregulated bi-directionally serving as the foundation of the dual algorithm development.

For the cluster analysis algorithm, the difference between Gene of Interest (GOI) CT values to GAPDH was normalized to the median of all samples for each given gene. In Genespring software, clustering was performed by hierarchical clustering design, with the Pearson-centered distance metric and centroid linkage rules; the distance metrics used here to assemble the

dendrogram are described in the paragraph after the next in details. Condition and gene entity are 2-D clustered in conjunction with expression heat map with red being up-regulation and blue being down-regulation.

For the ∑ΔCT algorithm, a dimensionality reduction formula was used to address the bi-directional signature in 1-D space (i.e. a number). The expression CT value of the housekeeping gene GAPDH was first subtracted from each EoE GOI CT value to acquire the ΔCT. The sums of the ΔCT were calculated separately for up-regulated and down-regulated gene groups. A negative weight was endowed to the up-regulated gene sum before the addition of the two ∑ΔCT values to establish the "EoE Score", reflecting the disease-specific expression signature and disease severity. The EoE R score (for remission and NL distinguishment) was calculated with the same formula considering the upregulated and downreguled genes following steroid treatment.

$$\begin{aligned}\text{EoE score} &= \sum \Delta CT \\ &= \sum_{i=1}^{27} (\text{CT down-regulated genes - CT GAPDH}) \\ &- \sum_{i=1}^{50} (\text{CT up-regulated genes - CT GAPDH})\end{aligned}$$

A distance metric calculation was also performed for clustering and dimensionality reduction (≥2-D). In this analysis, when sample similarity/dissimilarity were compared, two distance metrics were employed, namely Pearson-centered and Euclid-centered distances, for dendrogram assembling and multi-dimensional scaling (MDS) positioning, respectively. The formulas in the context of qPCR gene expression in "n-dimensional" space are given below. For any given pair of two samples (A, B) assayed with normalized CT values on an expression array of n

entities/genes, the n-dimensional EDP array data were represented as {A*i*, B*i*} (*i* = 1, 2, 3... n), where n = 77 (or other numbers applicable for distinct purposes) in the case of core EoE diagnostic genes.

$$\text{Pearson r} = \frac{\sum_{i=1}^{n} (A_i - \bar{A})(B_i - \bar{B})}{\sqrt{\sum_{i=1}^{n} (A_i - \bar{A})^2}\sqrt{\sum_{i=1}^{n} (B_i - \bar{B})^2}} \qquad \text{Euclid distance} = \sqrt{\sum_{i=1}^{n} (A_i - B_i)^2}$$

Supplemental References:

1. Blanchard C, Wang N, Stringer KF, et al. Eotaxin-3 and a uniquely conserved gene-expression profile in eosinophilic esophagitis. J Clin Invest 2006;116:536-47.
2. Caldwell JM, Blanchard C, Collins MH, et al. Glucocorticoid-regulated genes in eosinophilic esophagitis: a role for FKBP51. J Allergy Clin Immunol 2010;125:879-888 e8.
3. Abonia JP, Blanchard C, Butz BB, et al. Involvement of mast cells in eosinophilic esophagitis. J Allergy Clin Immunol 2010;126:140-9.
4. Straumann A, Bauer M, Fischer B, et al. Idiopathic eosinophilic esophagitis is associated with a T(H)2-type allergic inflammatory response. J Allergy Clin Immunol 2001;108:954-61.
5. Blanchard C, Mingler MK, McBride M, et al. Periostin facilitates eosinophil tissue infiltration in allergic lung and esophageal responses. Mucosal Immunol 2008;1:289-96.
6. Matoso A, Mukkada VA, Lu S, et al. Expression microarray analysis identifies novel epithelial-derived protein markers in eosinophilic esophagitis. Mod Pathol 2013;26:665-76.
7. Blanchard C, Stucke EM, Rodriguez-Jimenez B, et al. A striking local esophageal cytokine expression profile in eosinophilic esophagitis. J Allergy Clin Immunol 2011;127:208-17, 217 e1-7.

Table S1. The 96 genes embedded on the EDP with commercial ID number.

Gene	Fold change	Taqman Assay ID	Functionality
CDH26	27.2	Hs00375371_m1	Cell adhesion
CDH20	-3.6	Hs00230412_m1	Cell adhesion
CLDN10	-45.1	Hs01075312_m1	Cell adhesion
CTNNAL1	-7.8	Hs00972098_m1	Cell adhesion
DSG1	-153.1	Hs00355084_m1	Cell adhesion
CHL1	4.6	Hs00544069_m1	Cell adhesion
CXCL6	14.1	Hs00237017_m1	Chemokine
CCL26	194.7	Hs00171146_m1	Chemokine
CXCL1	64.4	Hs00236937_m1	Chemokine
IL8	27.0	Hs01553824_g1	Cytokine
IL5	3.7	Hs00174200_m1	Cytokine
IL13	19.9	Hs01124272_g1	Cytokine
CCR3	4.7	Hs99999027_s1	Eosinophilia
CLC	34.2	Hs00171342_m1	Eosinophilia
IL5RA	2.7	Hs00236871_m1	Eosinophilia
CRISP2	-31.6	Hs00162960_m1	Epithelial related
FLG	-10.0	Hs00863478_g1	Epithelial related
UPK1A	-21.9	Hs01086736_m1	Epithelial related
SPINK7	-12.6	Hs00261445_m1	Epithelial related
CRISP3	-205.5	Hs00195988_m1	Epithelial related
ACPP	-2.9	Hs00173475_m1	Epithelial related
UPK1B	5.1	Hs00199583_m1	Epithelial related
CA2	6.8	Hs00163869_m1	Epithelial related
PHLDB2	8.2	Hs00377503_m1	Epithelial related
MUC4	6.0	Hs00366414_m1	Epithelial related
GCNT3	3.1	Hs00191070_m1	Epithelial related
EPPK1	34.0	Hs02379935_s1	Epithelial related
ZNF365	-12.6	Hs00209000_m1	Inflammation process
CITED2	-4.0	Hs01897804_s1	Inflammation process
ARG1	-28.6	Hs00968979_m1	Inflammation process
ALOX12	-10.8	Hs00167524_m1	Inflammation process
IGJ	11.9	Hs00950678_g1	Inflammation process
TNFAIP6	389.3	Hs01113602_m1	Inflammation process
CFB	5.4	Hs00156060_m1	Inflammation process
HRH1	9.5	Hs00911670_s1	Inflammation process
CFI	5.1	Hs00989715_m1	Inflammation process
APOBEC3A	52.8	Hs00377444_m1	Inflammation process
MMP12	6.1	Hs00899668_m1	Inflammation process
CD200R1	27.3	Hs00708558_s1	Inflammation process
HPGDS	11.4	Hs00183950_m1	Inflammation process

FCGR3A/B	8.6	Hs00275547_m1	Inflammation process
RUNX2	6.4	Hs00298328_s1	Inflammation process
ALOX15	1110.5	Hs00609608_m1	Inflammation process
GRK5	5.2	Hs00992173_m1	Inflammation process
SAMSN1	8.7	Hs00223275_m1	Inflammation process
PMCH	127.0	Hs00173595_m1	Inflammation process
SLC16A6	-3.1	Hs00190779_m1	Ion channel
KCNJ2	5.2	Hs01876357_s1	Ion channel
ANO1	34.0	Hs00216121_m1	Ion channel
SLC26A4	27.4	Hs01070620_m1	Ion channel
TPSB2/AB1	9.9	Hs02576518_gH	Mastocytosis
CPA3	16.0	Hs00157019_m1	Mastocytosis
CMA1	5.1	Hs00156558_m1	Mastocytosis
NEFM	57.3	Hs00193572_m1	Neurosensory
NEFL	55.9	Hs00196245_m1	Neurosensory
PNLIPRP3	-15.3	Hs00406604_m1	Other
ENDOU	-22.9	Hs00195731_m1	Other
CDA	-56.1	Hs00156401_m1	Other
C7orf68	-16.8	Hs00203383_m1	Other
EML1	-15.0	Hs00270014_m1	Other
SUSD2	36.8	Hs00219684_m1	Other
GPR160	4.0	Hs01878570_s1	Other
TSPAN12	-7.1	Hs01113125_m1	Other
LRRC31	78.7	Hs00226845_m1	Other
GLDC	46.1	Hs01580586_g1	Other
GYS2	-24.5	Hs00608677_m1	Proliferation/growth
IGFL1	-18.4	Hs01651089_g1	Proliferation/growth
MT1M	-24.7	Hs00828387_g1	Proliferation/growth
CRYM	-44.4	Hs00157121_m1	Proliferation/growth
UBD	9.7	Hs00197374_m1	Proliferation/growth
GRPEL2	-4.7	Hs00537120_s1	Protein transport
RTP4	5.1	Hs00223342_m1	Protein transport
ACTG2	-19.7	Hs01123712_m1	Remodeling
CTSC	8.3	Hs00175188_m1	Remodeling
POSTN	48.0	Hs00170815_m1	Remodeling
KRT23	2.4	Hs00210096_m1	Remodeling
COL8A2	5.8	Hs00697025_m1	Remodeling
IL32	N/A	Hs00992441_m1	Cytokine
IL4	N/A	Hs99999030_m1	Cytokine
MSRB3	N/A	Hs00827017_m1	Other
CCL8	N/A	Hs00271615_m1	Chemokine
EPX	N/A	Hs00166795_m1	Eosinophilia
EPB41L3	N/A	Hs00202360_m1	Steroid-responding element

SYNPO2	**N/A**	**Hs02786674_s1**	**Other**
COL1A2	**N/A**	**Hs01028971_m1**	**Remodeling**
TRIM2	**N/A**	**Hs00209620_m1**	**Steroid-responding element**
18S	**N/A**	**Hs99999901_s1**	**Aux. housekeeping control**
SYNPO2L	**N/A**	**Hs00227561_m1**	**Other**
NCAM1	**N/A**	**Hs00287831_s1**	**Inflammation process**
F3	**N/A**	**Hs01076032_m1**	**Steroid-responding element**
TSLP	**N/A**	**Hs00263639_m1**	**Cytokine**
H19	**N/A**	**Hs00262142_g1**	**Steroid-responding element**
FKBP5	**N/A**	**Hs00296750_s1**	**Steroid-responding element**
SLAMF7	**N/A**	**Hs00900280_m1**	**Inflammation process**
PTGFRN	**N/A**	**Hs01385989_m1**	**Inflammation process**
GAPDH	**N/A**	**Hs03929097_g1**	**Main housekeeping control**

a. **The gene list presented herein is representative of the EDP design. Minor changes were made between different versions of the EDP without affecting the diagnostic merit, as the algorithm-based 77 diagnosis genes remained constant.**
b. **Fold change denotes level of dysregulation of NL vs. EoE, and N/A indicates non-applicable as no difference was present in EoE vs. NL diagnosis.**
c. **Assay ID denotes Taqman individual gene assay ID (Applied Biosystems)**

Table S2. The 22 steroid-responding genes distinguishing normal individuals from and patients with EoE in remission

Gene	Fluticasone[a]	Budesonide[a]	Taqman Assay ID
CRISP2	4.3	8.4	Hs00162960_m1
CHL1	6.9	4.1	Hs00544069_m1
GPR160	4.3	4.1	Hs01878570_s1
CFI	5.1	3.0	Hs00989715_m1
FKBP5	7.6	6.8	Hs00296750_s1
CTSC	2.7	2.5	Hs00175188_m1
CD200R1	57.3	57.0	Hs00708558_s1
TPSB2/AB1	2.4	2.4	Hs02576518_gH
SPINK7	2.7	3.5	Hs00261445_m1
CCL26	3.1	4.0	Hs00171146_m1
CRISP3	4.5	6.6	Hs00195988_m1
F3	2.5	3.1	Hs01076032_m1
CPA3	2.9	3.2	Hs00157019_m1
MUC4	6.5	5.7	Hs00366414_m1
EPB41L3	7.9	7.4	Hs00202360_m1
GCNT3	2.5	3.0	Hs00191070_m1
ACPP	2.0	2.1	Hs00173475_m1
CRYM	2.4	2.6	Hs00157121_m1
IL4	-14.6	-12.6	Hs99999030_m1
ACTG2	-24.7	-103.3	Hs01123712_m1
IL5RA	-9.8	-10.2	Hs00236871_m1
IL5	-8.2	-10.2	Hs00174200_m1

a. Fold change following steroid treatment, EoE Remission vs. Normal control

Table S3. Clinical information of subjects.

		Discovery cohorts		Replicating cohorts		Overall analysis	
		NL (n=14)	EoE (n=15)	Control (n=14)	EoE (n=18)	Control (n=50)	EoE (n=82)
Age mean		8.2	9.3	14.7	6.2	11.3y	9.3y
Age range		3y-19y	2y-17y	7y-58y	2y-13y	2y-59y	2y-45y
Gender (% Male)		57%	87%	64%	83%	74%	90%
Ethnity (% Caucacian)		93%	100%	93%	83%	94%	91%
Eosinophil/HPF mean		0.1	70	0.1	104	0.3	85
Eosinophil/HPF range		0-1	24-248	0-2	16-269	0-2	16-269
PPI confirmation*		78%	83%	67%	67%	76%	71%
History of EoE		0%	100%	29%	100%	50%	100%
History of EG**		7%	7%	7%	6%	10%	2%
History of other EGID		0%	0%	7%	0%	8%	1%
History of CE***		0%	0%	29%	0%	8%	2%
History of HES#		0%	0%	0%	0%	0%	0%
	Chronic Gastritis	7%	0%	21%	11%	14%	11%
	Chronic Duodenitis	7%	0%	7%	0%	6%	1%
History of other GI	Chronic Colitis	7%	0%	0%	0%	4%	0%
diagnostic abnormality	IBS	0%	0%	7%	0%	2%	0%
	Celiac disease	0%	0%	0%	0%	4%	0%
	GERD	0%	0%	0%	6%	2%	1%
	Marfan Syndrome	0%	0%	0%	0%	0%	1%
	Ehlers-Danlos Syndrome	0%	0%	0%	0%	0%	1%
Asthma		14%	27%	43%	33%	30%	35%
Allergic Rhinitis		36%	47%	64%	61%	58%	54%
Eczema		29%	27%	29%	44%	38%	41%
Urticaria		7%	0%	29%	44%	34%	40%
Swallowed glucocorticoid		0%	0%	14%	22%	38%	26%
Inhaled glucocorticoid		7%	13%	21%	6%	14%	5%
Ongoing diet treatment		N/A	20%	N/A	56%	N/A	44%

*Based on data available

**EG = Eosinophilic gastritis

***CE = Chronic esophagitis

HES = Hypereosinophilic syndrome

N/A = Not applicable